DØ Note 5288
November 22, 2006
Version 1.0

Using Bayesian Neural Networks to Search for Single Top Quarks in 1 fb^{-1} of Data

E. Aguiló,[2] P. Baringer,[10] A. Bean,[10] C. Belanger-Champagne,[11] J.A. Benitez,[12] E.E. Boos,[13] R. Brock,[12] V. Bunichev, [13] T. Burnett,[18] K. Chan,[2] L. Christofek,[4] J. Clutter,[9] Y. Coadou,[17] L.V. Dudko,[13] M. Erdmann,[1] T. Gadfort,[18] A. García-Bellido,[18] C. Gerber,[9] D. Gillberg,[17] G. Gutierrez,[7] P. Gutierrez,[14] A.P. Heinson,[5] U. Heintz,[3] S. Herrin,[16] S. Jabeen,[3] S. Jain,[14] A. Juste,[7] S. Kappler,[1] D. Kau,[8] G. Kertzscher,[11] M. Kirsch,[1] D. Lenz,[1] L. Li,[5] J. Mitrevski,[6] R. Moore,[2] M. Narain,[3] D. O'Neil,[17] G. Otero y Garzon,[9] M. Pangilinan,[3] J. Parsons,[6] M. Perfilov,[13] C. Potter,[11] H.B. Prosper,[8] R. Schwienhorst,[12] E. Shabalina,[9] J. Steggemann,[1] M. Strauss,[15] C. Tully,[14] M. Vetterli,[17] B. Vachon,[11] G. Watts,[18] M. Weber,[7]

[1] *RWTH Aachen University*
[2] *University of Alberta*
[3] *Boston University*
[4] *Brown University*
[5] *University of California, Riverside*
[6] *Columbia University*
[7] *Fermi National Accelerator Laboratory*
[8] *Florida State University*
[9] *University of Illinois, Chicago*
[10] *University of Kansas*
[11] *McGill University*
[12] *Michigan State University*
[13] *Moscow State University*
[14] *Princeton University*
[15] *University of Oklahoma*
[16] *Rice University*
[17] *Simon Fraser University*
[18] *University of Washington*

Using the selected event sets described in DØ Note 5285, we apply Bayesian neural networks to separate single top quark signals from the backgrounds, and use the resulting discriminant outputs to measure the cross sections for single top quark production for the first time.

Contents

1. INTRODUCTION

The top quark has been observed to decay into a W boson and a b quark. The same coupling must give rise to production of single top (or antitop) quarks via the electroweak interaction. The two leading contributions to single top production are $q\overline{q}' \rightarrow W^{+*} \rightarrow t\overline{b}$ (s-channel) or $qg \rightarrow qt\overline{b}$ (t-channel) - or charge conjugates. A measurement of the cross section for these processes would open a window into the direct study of the Wtb vertex with sensitivity to physics beyond the standard model.

Detecting single top production is difficult because not only is the production cross-section small (about 0.9 pb for the s-channel and 2.0 pb for the t-channel production) but the background is huge and very difficult to separate from signal. In order to reach the required sensitivity, we must either remove some of the background while keeping enough of the signal, or separate the signal from the background using carefully chosen variables which show discrimination between signal and background. With the usual approach of making cuts on these variables to improve the signal:background ratio, this is not achievable. We therefore have focus our efforts on identifying the best possible variables and developing multivariate methods using these variables for signal-background separation. There are many analysis techniques available that can be used for signal-background separation, for example, neural networks, decision trees, likelihood, matrix elements, etc. The method that we have decided to explore for this analysis is the use of Bayesian neural networks.

2. BAYESIAN NEURAL NETWORKS

Neural networks(NN) are a multivariate technique that is quite commonly used in high energy physics and they been effectively used previously in the Run I and the Run II analyses that searched for single top production.

Neural networks are parameterized nonlinear functions for regression or classification modeling. Inputs to a neural network are the variables that show discrimination between signal and background. Every network consists of at least three layers of nodes - an input layer, a hidden layer, and an output layer. At every hidden node, a sigmoid function is calculated from the sum of the weighted input variables. The linear sum of these sigmoid functions appears at the output node.

A neural network is trained with samples of simulated signal and background events. During the training process the weights are adjusted such that signal is moved towards 1 and background toward 0. We keep training until the maximal separation is achieved. A complete cycle of running through the entire training sample is called an "epoch". Many epochs may be required to achieve optimal training.

Although neural networks are popular both because of their power of classification and their ease of use, there are some limitations. The outcome of a standard training procedure is one point in the parameter space of the network weights, i.e. one network. One has to "decide" when to stop training, i.e. which network to use. Usually more complex models, that is networks with many nodes, give better results. However too complex a model is prone to overtraining. That is the result may be too specifically optimized to the training sample and generalize poorly to other samples. Thus large samples of training events are required.

The Bayesian approach averages over the outputs of many training epochs. Thus it is less prone to overtraining. One can choose very complex models without the danger of degradation in performance. The present analysis explores the use of Bayesian Neural Networks. We use the "Flexible Bayesian Modeling (FBM) package", by Radford Neal [1]. A detailed introduction of Bayesian Neural Networks and the use of this package can be

found in Ref. [1] and Ref. [2].

3. VARIABLE SELECTION

A list of sensitive variables has been derived based on an analysis of the signal and background Feynman diagrams [3, 4] and on a study of single top quark production at next-to-leading order [5]. The variables fall in three categories: individual object kinematics, global event kinematics, and variables based on angular correlations. The complete list of variables used in this analysis is shown in Table 1.

Bayesian NN variables
Object Kinematics
p_T(jet1)
p_T(best1)
p_T(tag1)
Event Kinematics
Aplanarity(alljets,W)
M(W,best1) ("best" top mass)
M(W,tag1) ("b-tagged" top mass)
M_T(W)
H_T(alljets)
H_T(alljets,W)
p_T(alljets−best1)
H_T(jet1,jet2)
H_T(jet1,jet2,W)
M(alljets−best1)
M(alljets−tag1)
p_T(alljets−best1)
M_T(jet1,jet2)
M_T(W)
Angular Correlations
cos(best1,notbest1)$_{\rm besttop}$
cos(tag1,lepton)$_{\rm btaggedtop}$
cos(lepton,besttopframe)$_{\rm besttopCMframe}$
cos(lepton,btaggedtopframe)$_{\rm btaggedtopCMframe}$
cos(lepton,Q(lepton)$\times z$)$_{\rm besttop}$
cos(notbest,alljets)$_{\rm alljets}$
$QTimesEta*$

TABLE 1: Variables used with the Bayesian NN, in three categories: object kinematics; event kinematics; and angular variables. For the angular variables, the subscript indicates the reference frame.

Previous iterations of the single top analysis at DØ have always considered fewer input variables. Introducing too many variables can degrade the performance of a network. We have not yet done specific studies to optimize the input variable and this is one of the main issues on our to do list. For now, we are using a combination of 24 variables listed in table 1. The same list is used in the electron and muon channel. Choice of 24 input variables is based on the fact that there is a cutoff at the number of variables that can be used for training in the FBM package implemented in DØ.

4. BAYESIAN NEURAL NETWORK TRAINING

The signal and background files were split into two different streams of samples. One was used for training the network and the other was used for measurement of yields. The split was done by examining the event number and if the event number was divisible by 3, then the event was used for training. All remaining events were used for measurements with their weights appropriately scaled.

Networks were trained for twelve separate channels, which were classified according to lepton flavor (electron and muon), the number of jets, and the number of b-tagged jets. We used events with two, three, or four jets and separate the samples further into samples with one or two b-tagged jets. For each channel, all background samples were combined, according to their weights, into a single background sample for training of the Bayesian Neural Network. Since our current modeling does not provide much statistics for the background from multijet events, we do not include this background for training purposes.

Training of the network was performed for the combined single top ($tb + tqb$) signal. We used the same filter functions to discriminate against all single top ($tb + tqb$), the s-channel (tb), and the t-channel (tqb) processes. The efficiency of these networks can bee seen from plots in Appendix 1. As can be seen from these plots, the discrimination is not yet optimal and in many channels, we still have some fraction of signal in lower discriminant region.

Each of the twelve networks was trained using 24 input variables and 40 hidden nodes. We run 800 iterations of training. Each iteration consists of 20 epochs and the result of each iteration is the average of the outputs of 20 training epochs. The final network for each one of the twelve channels is averaged over the last 100 iterations.

5. CROSS CHECK SAMPLES

In order to validate every step of the Bayesian neural network analysis cross check samples were used to decide whether the background model and data are in agreement after applying a Bayesian neural network. Two cross check samples are defined (see main single top note for details) as follows: "W+jets" (2 jets, 1 b-tag, $H_T < 175$ GeV), and "$t\bar{t}$" (4 jets, 1 b-tag, $H_T > 300$ GeV). These samples are designed such that on of them has mostly W+jets and almost no ttbar, and the other is mostly ttbar and almost no W+jets. This allows us to test whether or not each part of the background model is adequately described. Figure 1 shows the Bayesian neural network output distributions for these cross check samples in the *tbtqb* channel. More plots are available in Appendix 1:

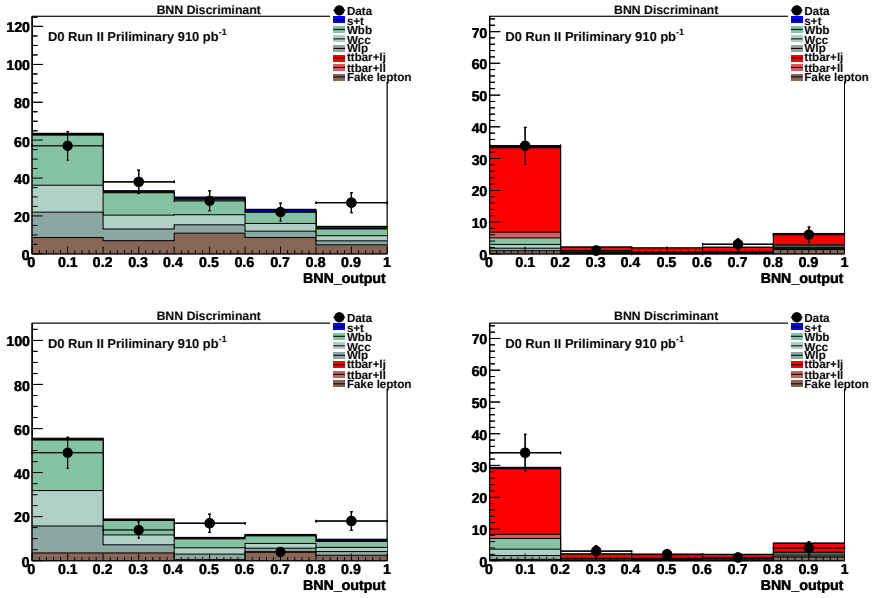

FIG. 1: Bayesian neural network outputs for the cross check samples in the *tbtqb* channel. Top row: electron channel; bottom row: muon channel. Left column: "W+jets" sample; right column: "$t\bar{t}$" sample.

From the information in all such histograms we concluded there is no obvious bias in our measurement. The background model describes the data within uncertainties.

6. ENSEMBLE TESTS

Several subsets of psuedo-data events have been produced from the background model to test the performance of the analysis. In each case many fake datasets were generated. These ensembles include:

1. 1000 experiments with SM single top content (2.9 pb)
2. 4 sets of 100 experiments with unknown (to the analyzers) single top content, but SM ratios between the tb and tqb-channels.
3. 1000 experiments with NO single top signal, only backgrounds.

The full analysis chain was run using each fake-data set as if it were the real one, and the results are presented in the following subsections.

6.1. Results of the Ensemble Tests

SM Ensembles:
Figure 2 shows the cross sections calculated after applying the Bayesian neural networks to the 1000 ensembles with SM signal content. The mean of each of the distributions in Fig. 2 is shifted slightly with respect to the input value. For example, the mean, with full systematics is 3.12 pb in tb+tqb for an input of 2.90 pb. This shift is consistent with what is expected from the standard model input of 2.90 pb for tb+tqb sample. For the no systematics case the mean is 2.84 pb.

The Zero Signal Ensembles Set:
The mean of each of the distributions in Fig. 3 is consistent with zero, as expected.

The Non-SM Ensembles with SM tb:tqb Ratio
Here we show the tb+tqb result for ensembles with different SM tb:tqb ratios. Results are shown in Fig. 4 for the four ensembles of this type, both with and without the full systematic uncertainties included.

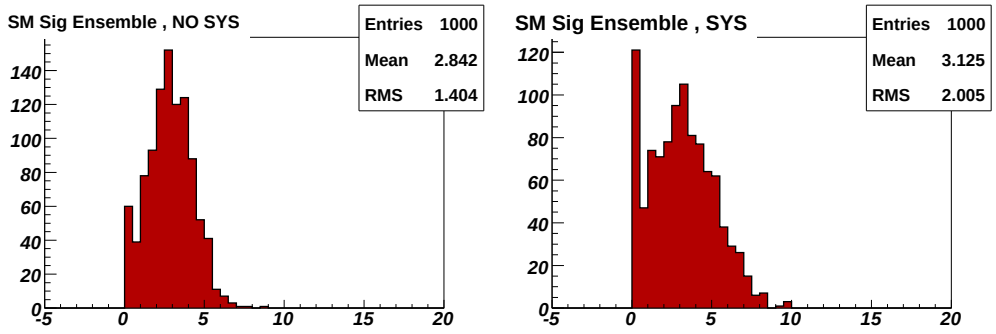

FIG. 2: Results using the 1000 ensembles with SM cross sections for the tb+tqb channel results, with no systematic uncertainties included (left), and with systematic uncertainties included (right).

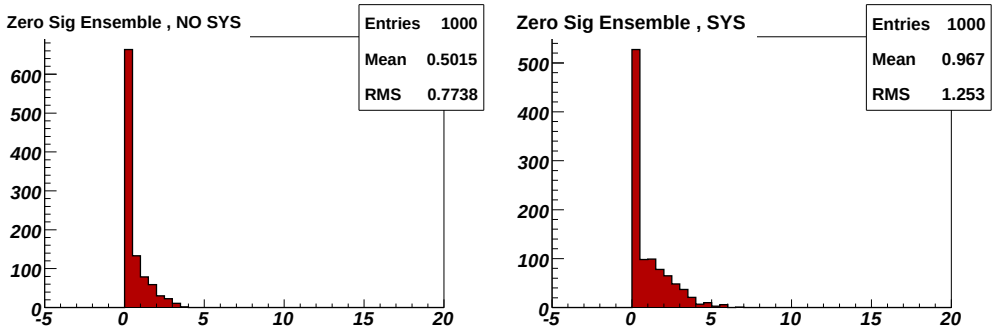

FIG. 3: Results using the 1000 ensembles with zero tb:tqb channel signal cross sections sections for the tb+tqb channel results, with no systematic uncertainties includeded (left), and with systematic uncertainties included (right).

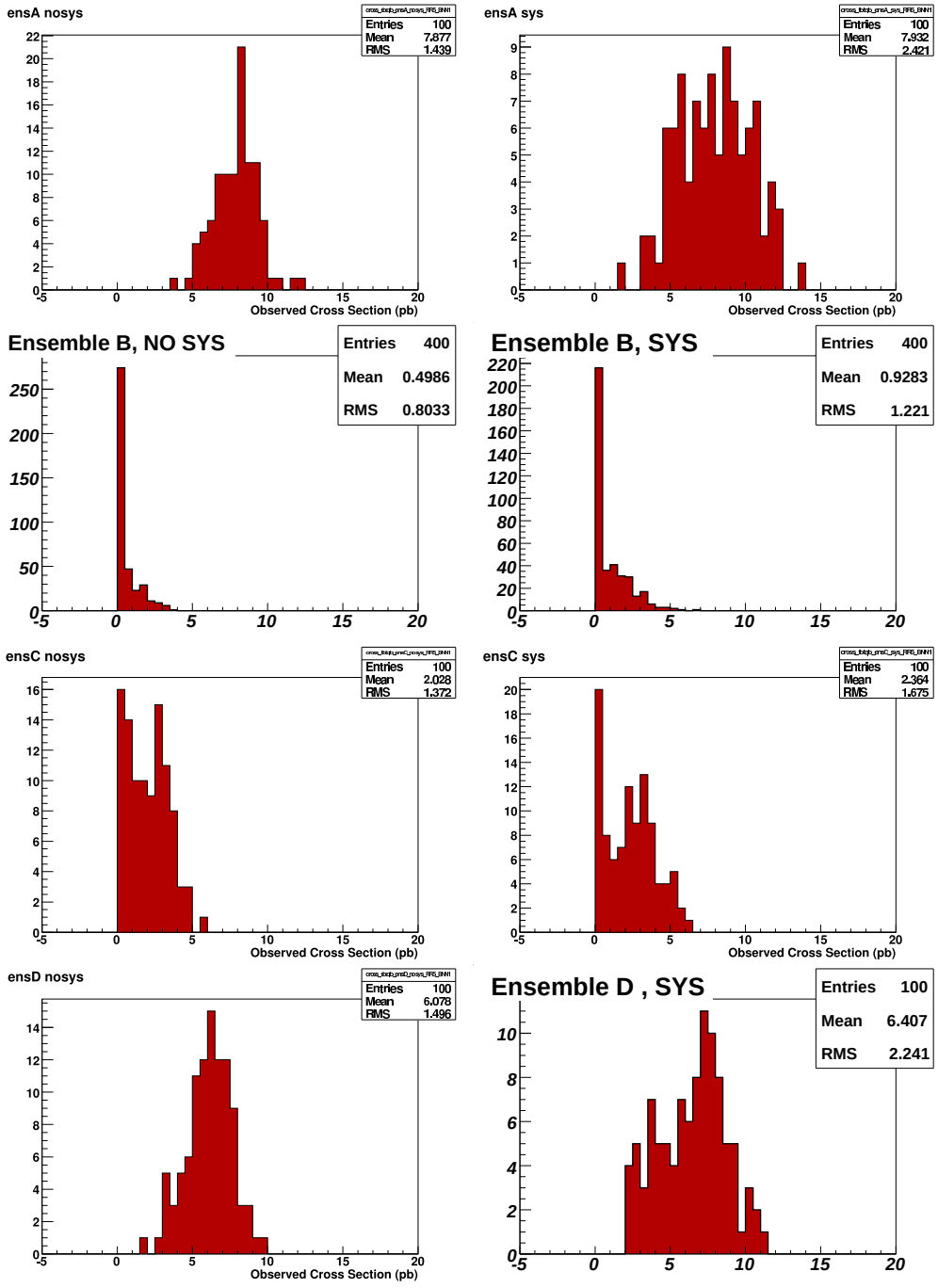

FIG. 4: Results using the ensembles with non-SM cross section but SM *tb*:*tqb* ratio. The left column contains tests without systematic errors, while the right column includes the full list of systematic errors. The top row contains ensembles A, the second and the third row are ensembles B, and C respectively, and the lowest row shows results from ensemble D.

Our measured values of the cross sections are:

- A: 2.54 × SM (true value = 2.76 × SM)
- B: 0.97 (true value = 0)
- C: 0.76 × SM (true value = 0.70 × SM)
- D: 2.06 × SM (true value = 2.00 × SM)

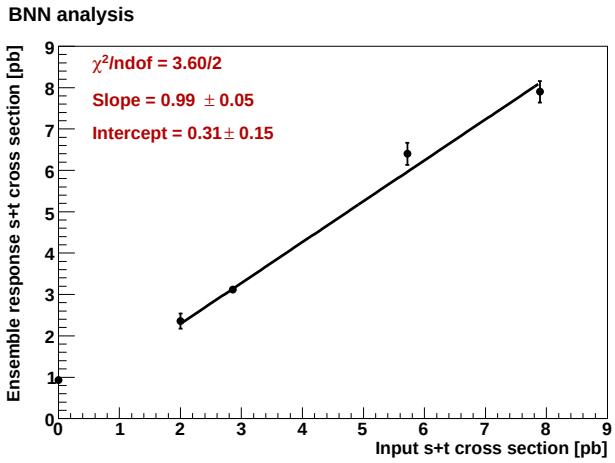

FIG. 5: Correlation between input tb:tqb cross section and response of non-SM ensembles.

6.2. Effect of Systematics from SM ensemble tests

We used 200 of the 1000 set SM ensembles to perform a study of the effects of various individual systematic uncertainties on the cross section measurement. These tests are shown in Fig. 6 through Fig. 8. We find that binned uncertainties, jet energy scale and b-tagging TRFs are rather well behaved. The largest broadening comes from the combined flat systematic errors. Within these flat systematic errors, the biggest source of uncertainty arises from the normalization of the W+jets and multijets events. and then the top production cross section. These two sources tend to give large RMS spread and since the cross section is bounded and is not allowed to go below zero, this broadening gives rise to a peak at zero cross section.

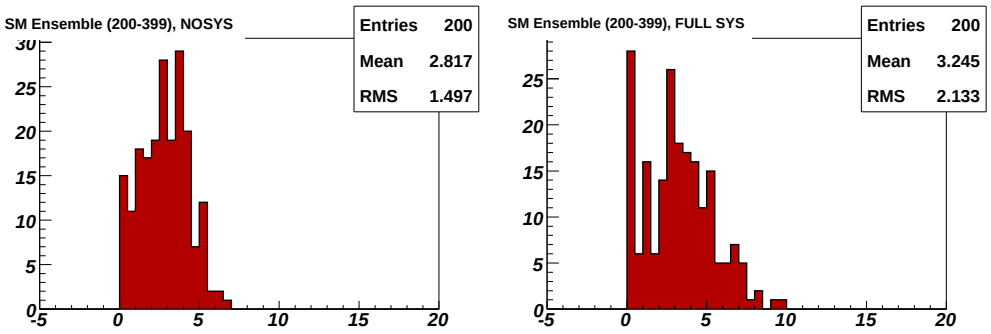

FIG. 6: Results of applying no systematic uncertainty (left plot) and full systematic uncertainty (right plot) to 200 of the SM ensembles.

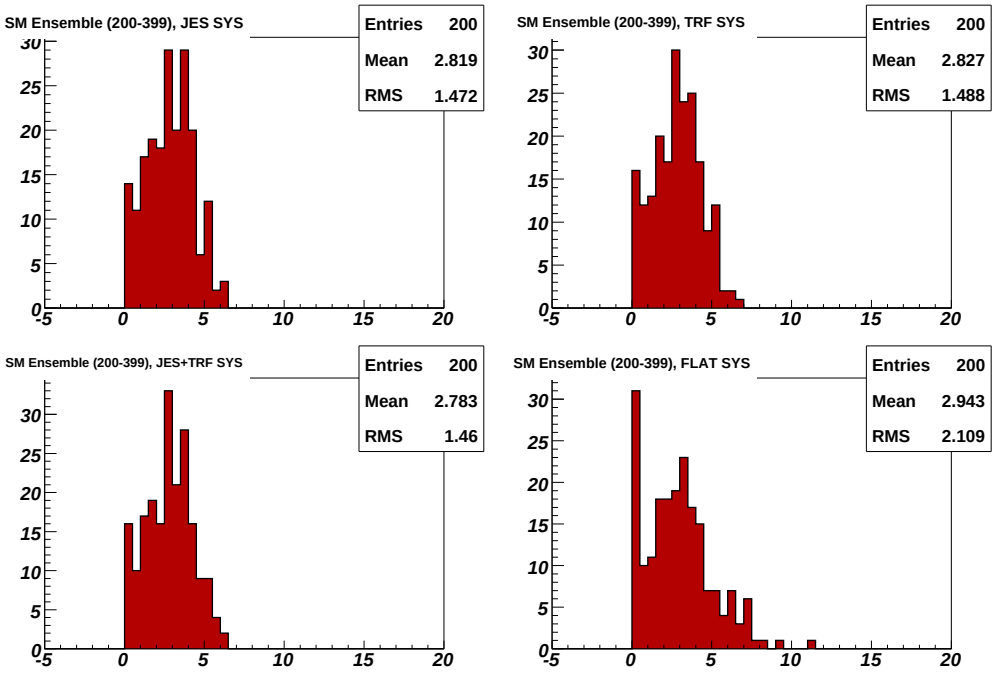

FIG. 7: Results of applying binned JES systematic uncertainty (upper left plot), the binned TRF systematic uncertainty (upper right plot). The effect of the combined JES and TRF uncertainty and the effect of the rest of the uncertainties are shown in the lower left and right plots respectively. Here 200 of the SM ensembles are used for the study.

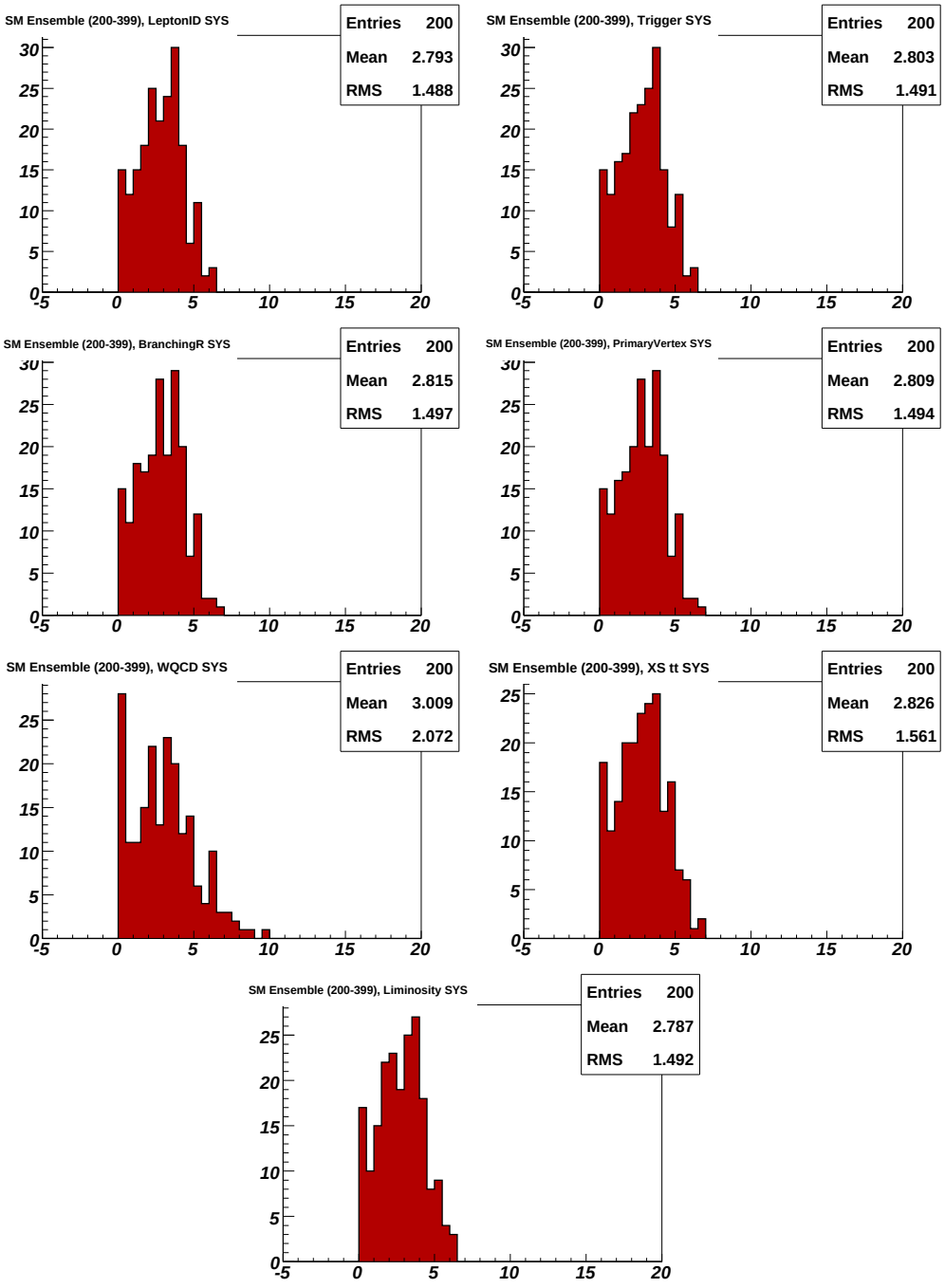

FIG. 8: Results of applying some of the components of the systematic uncertainties to 200 of the SM ensembles.

7. EXPECTED RESULTS

Since the agreement is good between the data and background model in the Bayesian neural network output for the control samples, we proceed to make a measurement. In order to decide which analysis channels to combine to get the best expected sensitivity, we calculate for each channel and in all combinations the expected Bayes ratio and we check the result by calculating also the expected cross section uncertainty.

The Bayes ratios for various combinations of channels are reported in Table 2. A ratio of 1.0 means that there is no signal seen in that channel. The larger the Bayes ratio value the better, although once the posterior density distribution is Gaussian (i.e., a clear signal is observed), then arbitrarily large values are obtained and the size ceases to be useful for comparing channels or analyses.

Expected Bayes Ratios

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Statistics only								
tb	1.3	1.1	1.2	1.2	1.3	1.1	1.0	**1.5**
tqb	1.9	1.4	2.6	1.0	2.1	1.2	1.0	**2.7**
tbtqb	4.0	2.3	6.8	1.4	5.2	1.7	1.1	**9.2**
With systematics								
tb	1.1	1.0	1.1	1.0	1.2	1.0	1.0	**1.3**
tqb	1.3	1.3	2.4	1.0	2.1	1.0	1.0	**2.7**
tbtqb	1.6	1.2	3.5	1.1	2.8	1.1	1.0	**4.5**

TABLE 2: Expected Bayes ratios, without and with systematic uncertainties, for many combinations of analysis channels. The best values from all channels combined, with systematics, are shown in bold type.

Table 3 shows another parameter useful for comparing the sensitivity of different analysis channels, the half the full width of the expected cross section posterior density distribution, divided by the cross section at the peak, which gives approximately the cross section uncertainty. The smaller the value the better the expected measurement.

Relative Uncertainties on the Expected Cross Sectionsin

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Statistics only								
tb	101.7 %	130.0 %	110.0 %	116.7 %	95.0 %	167.2 %	494.8 %	88.3 %
tqb	79.5 %	92.4 %	67.4 %	275.0 %	74.2 %	109.1 %	280.3 %	66.7 %
tbtqb	58.9 %	71.6 %	51.6 %	101.6 %	55.2 %	83.2 %	193.7 %	48.4 %
With systematics								
tb	212.5 %	–	175.9 %	234.1 %	130.0 %	5025.0 %	–	**107.4 %**
tqb	108.8 %	115.3 %	75.6 %	–	78.2 %	224.5 %	2515.4 %	**71.2 %**
tbtqb	90.4 %	129.6 %	67.0 %	217.2 %	68.3 %	179.3 %	1002.9 %	**60.7 %**

TABLE 3: Relative uncertainties on the expected cross sections, without and with systematic uncertainties, for many combinations of the analysis channels.

It can been seen from the numbers in Tables 2 and 3 that combining the two single top signals (tb+tqb) gets the best expected sensitivity. The single-tag two-jet channel contributes most to this sensitivity, as expected from the high signal acceptance and reasonable signal:background ratio.

Table 4 shows the expected cross section results for the same subsets of combined channels.

Expected Single Top Cross Sections [pb]								
	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Statistics only								
tb	$0.9^{+0.9}_{-0.9}$	$0.9^{+1.4}_{-0.9}$	$0.9^{+1.1}_{-0.9}$	$0.9^{+1.2}_{-0.9}$	$0.9^{+0.9}_{-0.8}$	$0.9^{+2.0}_{-0.9}$	$0.9^{+7.7}_{-0.9}$	$0.9^{+0.8}_{-0.8}$
tqb	$2.0^{+1.6}_{-1.5}$	$2.0^{+1.9}_{-1.8}$	$2.0^{+1.4}_{-1.3}$	$2.0^{+8.9}_{-2.0}$	$2.0^{+1.5}_{-1.4}$	$2.0^{+2.3}_{-2.0}$	$2.0^{+9.1}_{-2.0}$	$2.0^{+1.4}_{-1.3}$
$tbtqb$	$2.9^{+1.7}_{-1.6}$	$2.8^{+2.1}_{-2.0}$	$2.9^{+1.5}_{-1.5}$	$2.8^{+3.0}_{-2.8}$	$2.9^{+1.6}_{-1.6}$	$2.8^{+2.4}_{-2.3}$	$2.8^{+8.2}_{-2.8}$	$2.9^{+1.4}_{-1.4}$
With systematics								
tb	$0.6^{+2.0}_{-0.6}$	$0.0^{+3.4}_{-0.0}$	$0.9^{+2.2}_{-0.9}$	$0.7^{+2.4}_{-0.7}$	$0.9^{+1.4}_{-0.9}$	$0.1^{+6.0}_{-0.1}$	$0.0^{+18.6}_{-0.0}$	$\mathbf{1.0^{+1.2}_{-1.0}}$
tqb	$1.7^{+2.0}_{-1.7}$	$2.2^{+2.8}_{-2.2}$	$2.3^{+1.9}_{-1.6}$	$0.0^{+16.6}_{-0.0}$	$2.3^{+1.9}_{-1.7}$	$1.5^{+5.1}_{-1.5}$	$0.4^{+19.2}_{-0.4}$	$\mathbf{2.4^{+1.8}_{-1.6}}$
$tbtqb$	$2.5^{+2.4}_{-2.1}$	$2.1^{+3.4}_{-2.1}$	$3.1^{+2.2}_{-1.9}$	$2.0^{+6.7}_{-2.0}$	$3.1^{+2.2}_{-2.0}$	$2.1^{+5.4}_{-2.1}$	$1.0^{+19.4}_{-1.0}$	$\mathbf{3.2^{+2.1}_{-1.8}}$

TABLE 4: Expected cross sections, without and with systematic uncertainties, for many combinations of the analysis channels. The final expected results of this analysis are shown in the lower right hand corner in bold type.

The posterior distributions for the final combinations with systematics are shown in Fig. 9. These plots represent the final expected results from this analysis.

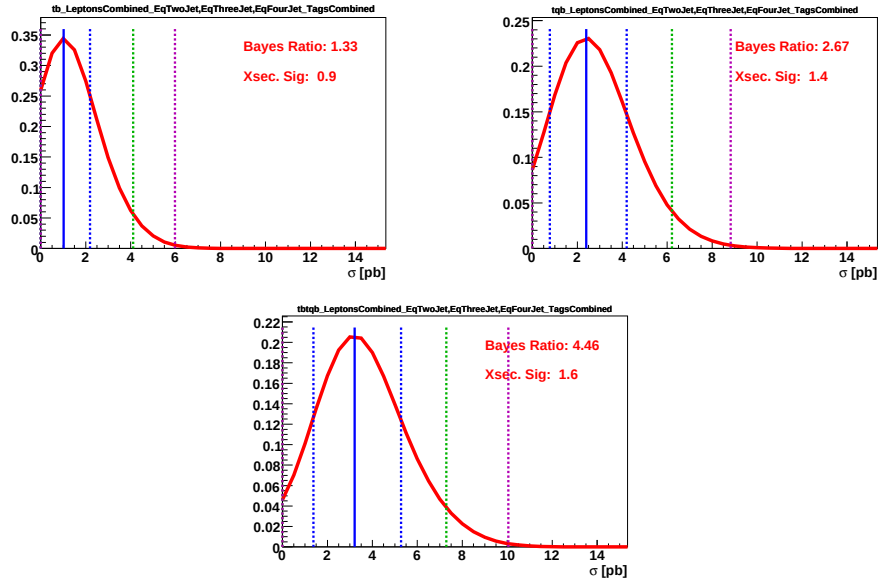

FIG. 9: Expected posterior density distributions from Bayesian neural network for s-channel tb (top left plot), t-channel tqb (top right plot) and $s+t$-channel $tbtqb$ (lower plot) as signal, for lepton flavor (e,μ), number of b-tagged jets (1,2), and jet multiplicity (2,3,4) combined, with all systematic uncertainties included.

8. OBSERVED RESULTS

This section contains Bayesian neural network observed cross section results using the $\sim$1 fb^{-1} dataset. We show in Fig. 10 the Bayesian neural network outputs from the two most sensitive channels, those with two jets with one of them tagged. The figures in Appendix 2 show the Bayesian neural network outputs for all channels separately.

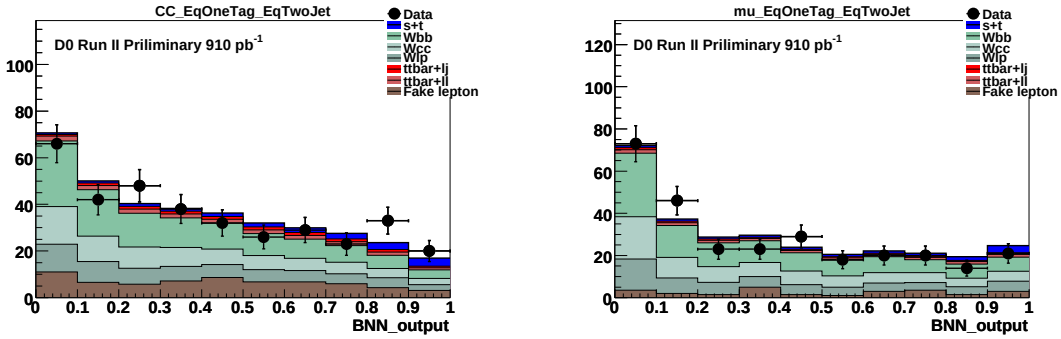

FIG. 10: Output distributions from the Bayesian neural networks trained to find the summed tb+tqb signal. The left plot is the electron channel and the right plot for the muon channel. Both channels use events with two jets, one of them tagged, which is the most sensitive of all the channels.

Table 5 shows the measured cross sections from various combinations of analysis channels.

Measured Single Top Cross Sections [pb]								
	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Statistics only								
tb	$2.1^{+1.3}_{-1.1}$	$1.4^{+1.3}_{-1.3}$	$1.0^{+1.0}_{-1.0}$	$2.8^{+1.6}_{-1.4}$	$2.2^{+1.2}_{-1.1}$	$0.7^{+2.0}_{-0.7}$	$0.0^{+7.5}_{-0.0}$	$1.9^{+1.0}_{-1.0}$
tqb	$3.2^{+1.8}_{-1.7}$	$0.3^{+2.3}_{-0.3}$	$1.9^{+1.3}_{-1.2}$	$12.8^{+9.2}_{-8.6}$	$3.3^{+1.7}_{-1.6}$	$0.0^{+2.5}_{-0.0}$	$0.0^{+9.8}_{-0.0}$	$2.2^{+1.4}_{-1.3}$
$tbtqb$	$3.2^{+1.7}_{-1.6}$	$0.9^{+1.9}_{-0.9}$	$1.7^{+1.3}_{-1.2}$	$7.0^{+4.0}_{-3.7}$	$3.4^{+1.6}_{-1.5}$	$0.0^{+2.6}_{-0.0}$	$0.0^{+9.1}_{-0.0}$	$2.4^{+1.3}_{-1.3}$
With systematics								
tb	$4.7^{+2.5}_{-2.1}$	$3.7^{+1.9}_{-2.8}$	$3.5^{+1.9}_{-2.2}$	$4.4^{+3.2}_{-2.5}$	$3.7^{+1.7}_{-1.7}$	$2.3^{+4.5}_{-2.3}$	$0.0^{+19.5}_{-0.0}$	$\mathbf{3.8^{+1.4}_{-1.5}}$
tqb	$5.3^{+2.7}_{-2.5}$	$0.6^{+3.8}_{-0.6}$	$3.8^{+1.9}_{-1.8}$	$15.9^{+18.0}_{-13.3}$	$4.4^{+2.3}_{-2.0}$	$0.0^{+4.6}_{-0.0}$	$2.6^{+18.9}_{-2.6}$	$\mathbf{3.7^{+2.0}_{-1.8}}$
$tbtqb$	$6.7^{+2.9}_{-2.7}$	$2.2^{+2.7}_{-2.2}$	$4.0^{+1.9}_{-2.0}$	$13.0^{+9.0}_{-7.2}$	$5.5^{+2.2}_{-2.2}$	$0.4^{+5.2}_{-0.4}$	$1.7^{+19.8}_{-1.7}$	$\mathbf{5.0^{+1.9}_{-2.0}}$

TABLE 5: Measured cross sections, without and with systematic uncertainties, for many combinations of the analysis channels. The final expected results of this analysis are shown in the lower right hand corner in bold type.

Table 6 shows the total uncertainties on the cross sections as percentages. These values are calculated as the difference between the peak value plus one sigma and minus one sigma, divided by two times the peak value.

Relative Uncertainties on the Measured Cross Sections

	1,2tags + 2,3,4jets		e,μ + 2,3,4jets		e,μ + 1,2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
Statistics only								
tb	58.0 %	94.6 %	101.5 %	54.9 %	52.1 %	204.5 %	–	52.4 %
tqb	55.2 %	430.0 %	66.4 %	70.0 %	50.0 %	–	–	60.4 %
tbtqb	51.9 %	155.0 %	71.9 %	55.0 %	46.1 %	–	–	53.1 %
With systematics								
tb	49.4 %	62.5 %	58.2 %	64.7 %	46.3 %	146.2 %	–	**39.0 %**
tqb	49.2 %	386.8 %	48.4 %	98.2 %	49.0 %	–	408.0 %	**51.2 %**
tbtqb	41.9 %	110.0 %	48.9 %	62.2 %	40.7 %	626.7 %	628.9 %	**38.6 %**

TABLE 6: Relative uncertainties on the measured cross sections, without and with systematic uncertainties, for many combinations of the analysis channels.

The posterior distributions for the final combinations with systematics are shown in Fig. 11. These plots represent the final measured results from this analysis.

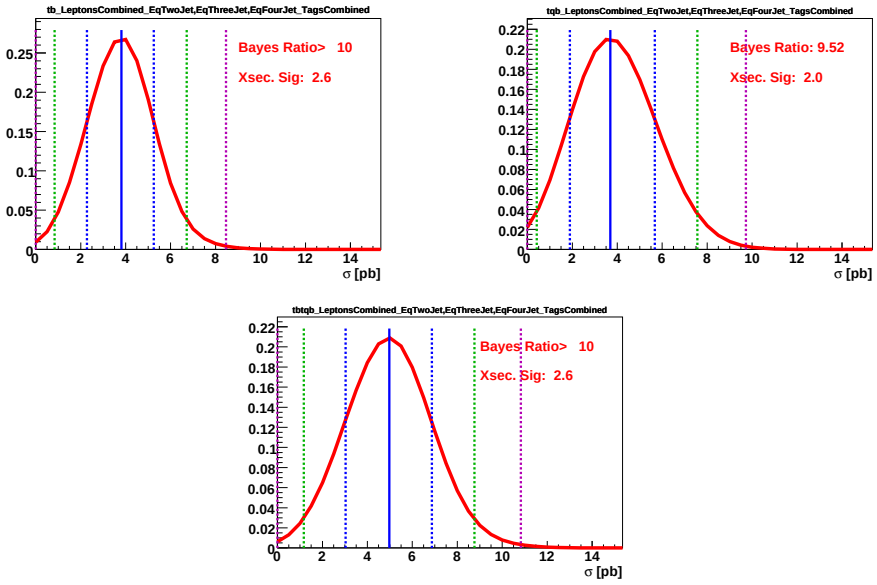

FIG. 11: Observed posterior density distributions from Bayesian neural network outputs for s-channel tb (top left plot), t-channel tqb (top right plot) and $s+t$-channel $tbtqb$ (lower plot) as signal, for lepton flavor (e,μ), number of b-tagged jets (1,2), and jet multiplicity (2,3,4) combined, with all systematic uncertainties included.

Figure 12 shows the cross sections measured for combined tb+tqb production in each independent analysis channel, and the combined result, taken from the 1-d posterior density distribution measurements.

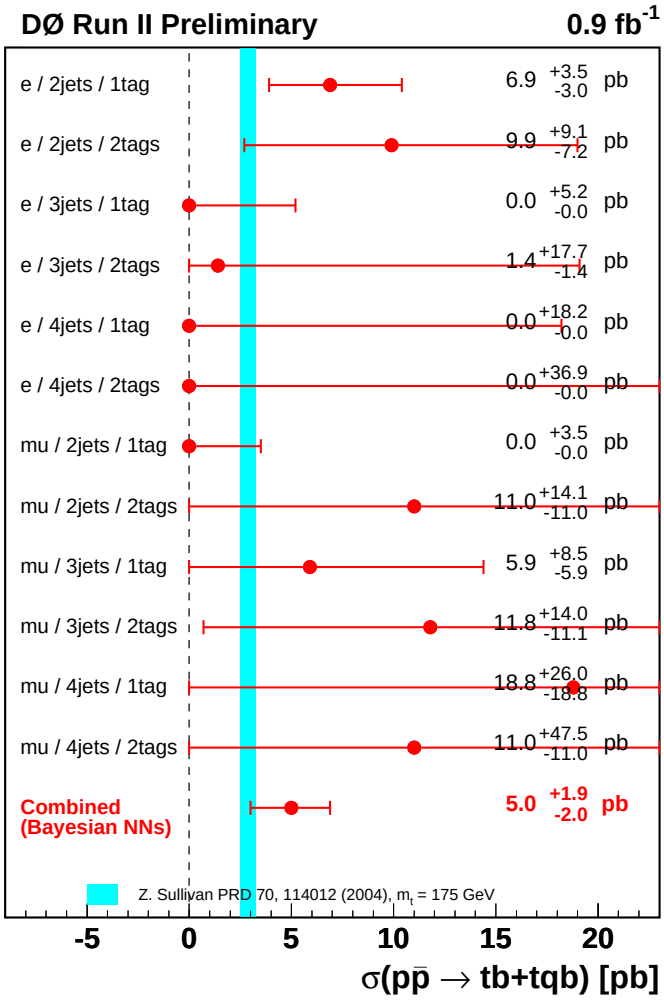

FIG. 12: Summary plot of the measured single top quark cross sections showing the individual measurements and their combination.

9. EVENT CHARACTERISTICS

We would like to know if the events with high discriminant value look top-quark-like. If they do, it is not proof we have observed single top quark production, but if they do not, it would be problematic. Therefore, we make a high cut (> 0.98) on the discriminant output distributions from each analysis channel and compare it with background-like data events that have a discriminant cut of < 0.6. These events cluster around $M_T(W) = 75$ GeV because they contain real W bosons, but do not exhibit the strong peak around $M(W, \text{tag1}) = 175$ GeV.

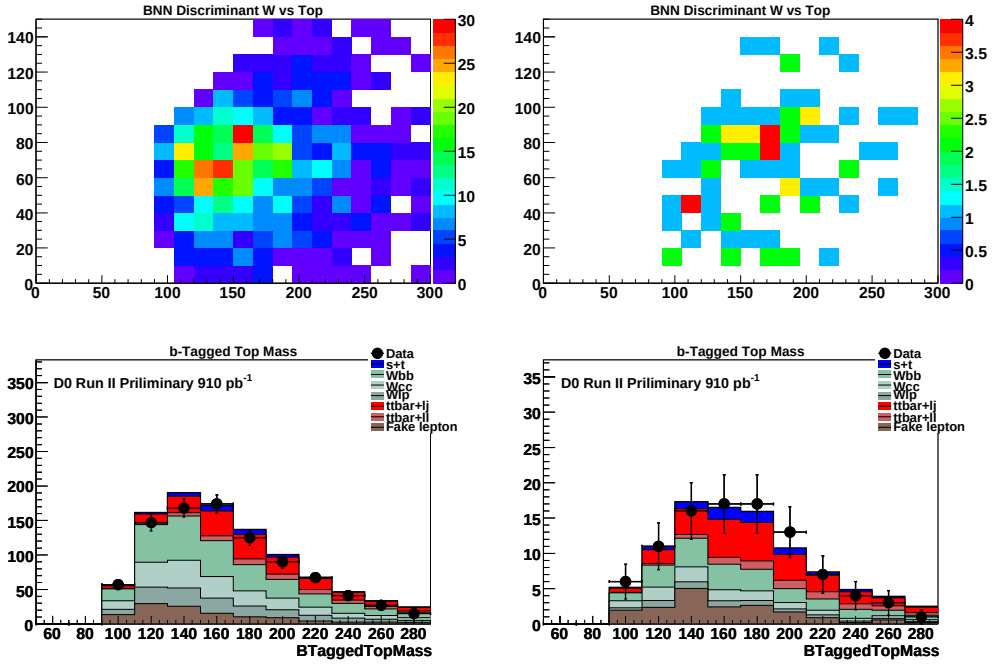

FIG. 13: Upper row: W transverse mass vs b-tagged top quark mass in the $tbtqb$ channels for events with a low BNN output < 0.6 (left column) and high BNN output > 0.98 (right column), for electron and muon, one and two b-tagged jets and two, three, and four jets channels combined; Lower row: b tagged top mass data vs background sum plots for same cuts.

10. SUMMARY

We have used Bayesian neural networks to separate expected single top quark signals from background, and used the resulting discriminant outputs to measure the single top cross sections. We obtain the following results: For 2,3,4 jets combined

$$\begin{aligned} &\text{s-channel :} && \sigma\,(p\bar{p} \to tb + X) && = 3.8^{+1.4}_{-1.5}\ \text{pb} \\ &\text{t-channel :} && \sigma\,(p\bar{p} \to tqb + XX) && = 3.7^{+2.0}_{-1.8}\ \text{pb} \\ &\text{s + t -channels :} && \sigma\,(p\bar{p} \to tb + tqb + X) && = 5.0^{+1.9}_{-2.0}\ \text{pb} \end{aligned}$$

11. COMMENTS

This note is not yet complete. Things we are working on include:

- Calculate p-values (significances) for the results.

This is the first time that BNN is being used in a full analysis in DØ. There are many aspects of the analysis that can be improved for example to find a suitable set of optimized input variables. Unlike some other multivariate techniques, irrelevant variables may deteriorate the performance of a network. Exploring methods of finding relevant variables, is our highest priority for the near future BNN analysis.

[1] http://www.cs.toronto.edu/radford/fbm.software.html
[2] Suman B. Beri, Rajwant Kaur, Harrison B. Prosper, "Bayesian Neural Networks," D O note 4846.
[3] L. Dudko, "Use of Neural Networks in a Search for Single Top Quark Production at DØ," AIP Conf. Proc. **583**, 83 (2001).
[4] E. Boos and L. Dudko, "Optimized Neural Networks to Search for Higgs Boson Production at the Tevatron," Nucl. Instrum. Meth. in Phys. Res. **A 502**, 486 (2003).
[5] Q.H. Cao, R. Schwienhorst, and C.P. Yuan, "Next-to-Leading Order Corrections to Single Top Quark Production and Decay at the Tevatron: 1: s-Channel Process," Phys. Rev. D **71**, 054023 (2005).

APPENDIX 1 — BAYESIAN NEURAL NETWORK EFFICIENCY

NORMALIZED BAYESIAN NEURAL NETWORK OUTPUTS AND EFFICIENCIES FOR THE *tb+tqb* ELECTRON ONE TAG ANALYS

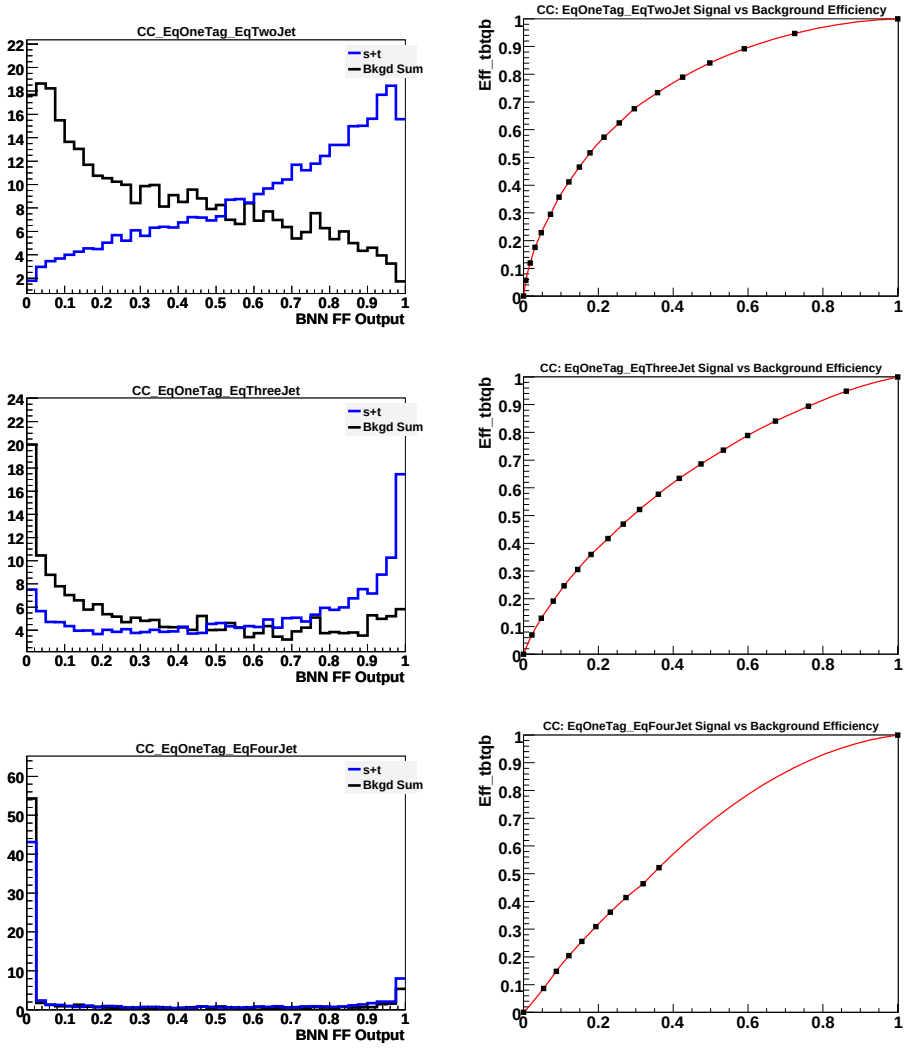

FIG. 14: Left column: normalized Bayesian neural network output for the *tbtqb* signal in the electron channel for events with one tagged jet. Left column: corresponding efficiency of the network, where the x-axis is the efficiency for the background and the y-axis is the efficiency for the signal.

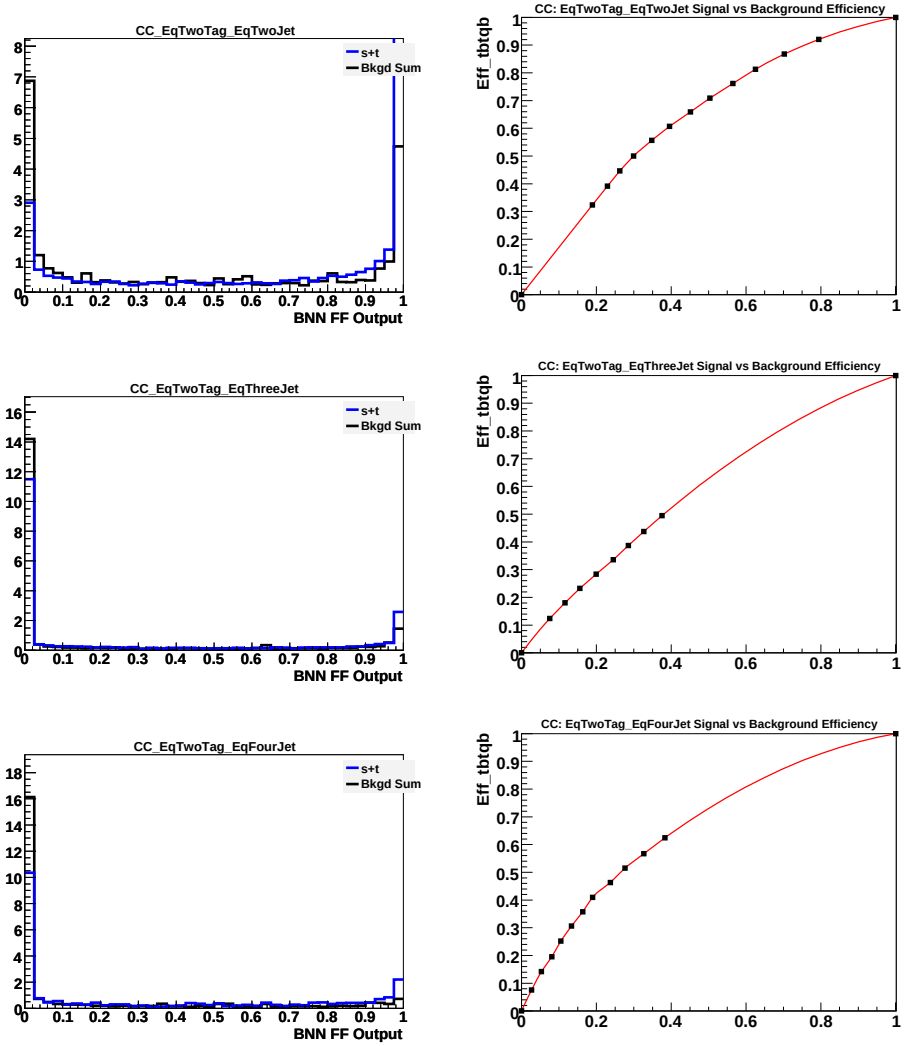

FIG. 15: Left column: normalized Bayesian neural network output for the *tbtqb* signal in the electron channel for events with two tagged jets. Left column: corresponding efficiency of the network, where the x-axis is the efficiency for the background and the y-axis is the efficiency for the signal.

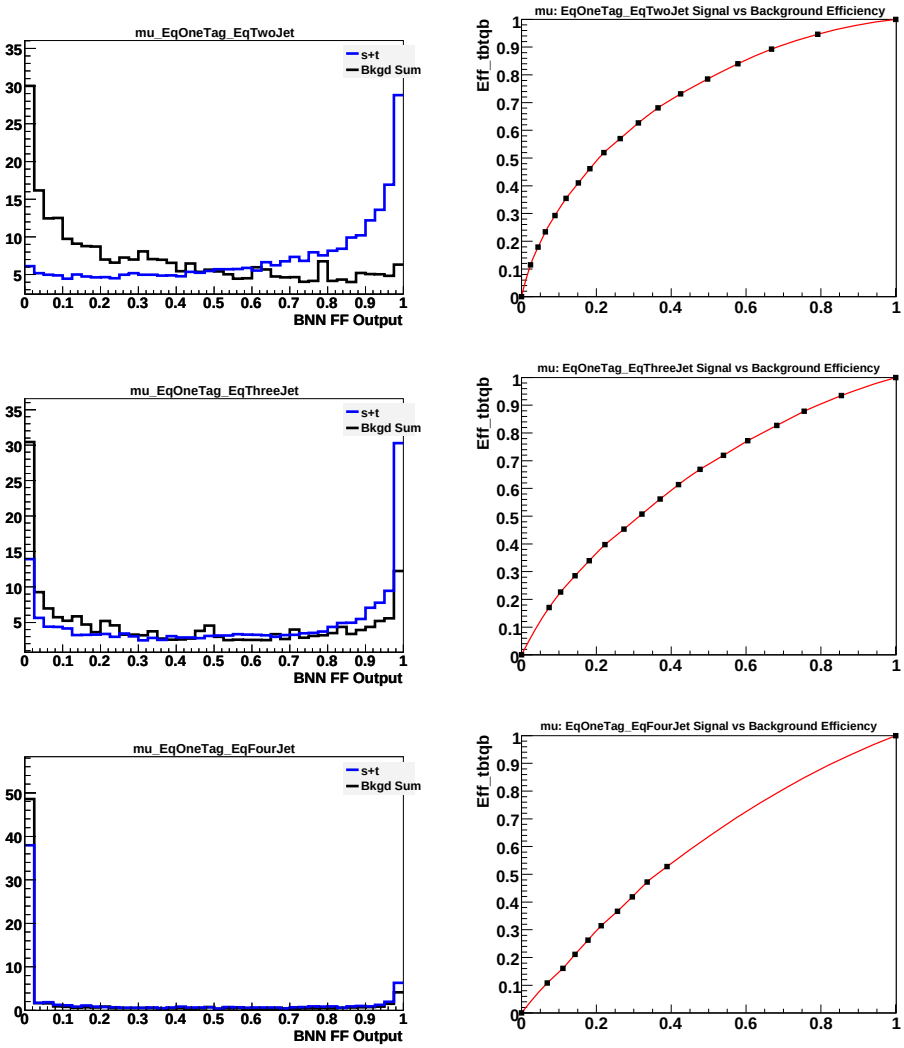

FIG. 16: Left column: normalized Bayesian neural network output for the *tbtqb* signal in the muon channel for events with one tagged jet. Left column: corresponding efficiency of the network, where the x-axis is the efficiency for the background and the y-axis is the efficiency for the signal.

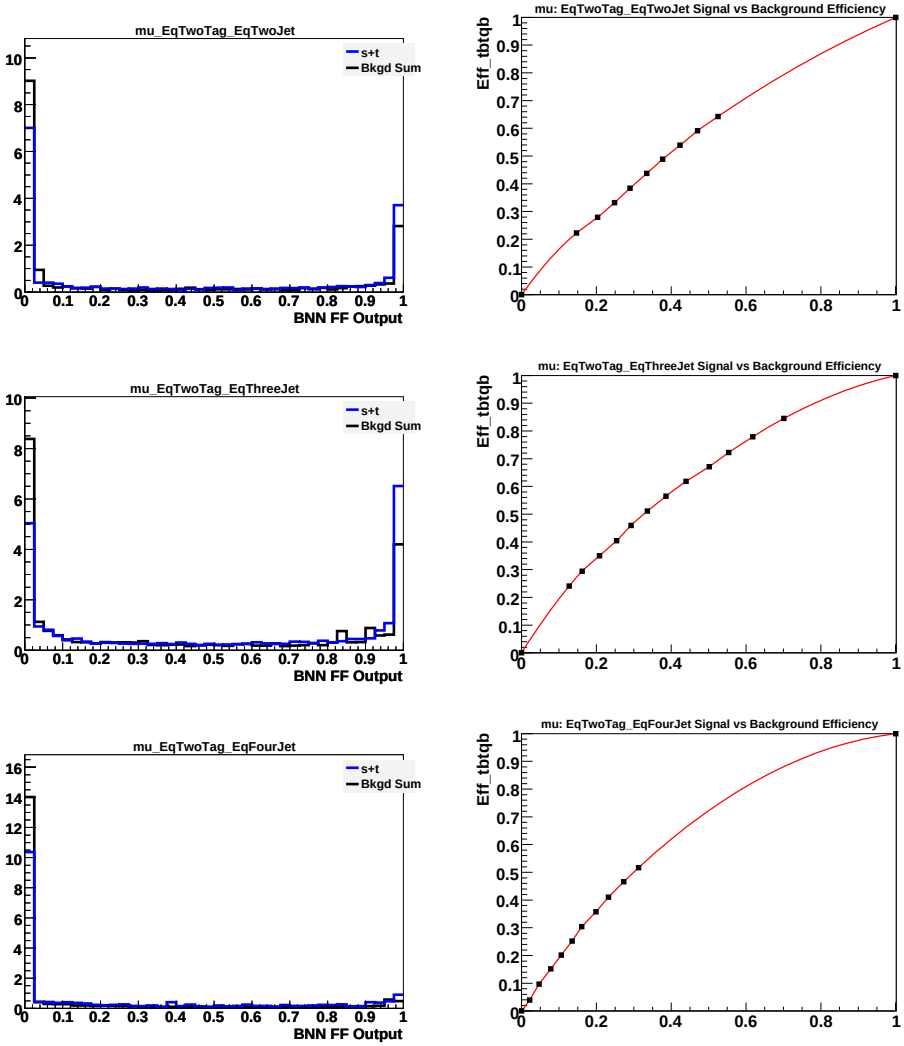

FIG. 17: Left column: normalized Bayesian neural network output for the $tbtqb$ signal in the muon channel for events with one tagged jet. Left column: corresponding efficiency of the network, where the x-axis is the efficiency for the background and the y-axis is the efficiency for the signal.

APPENDIX 2 — BAYESIAN NEURAL NETWORK OUTPUT PLOTS FROM ANALYSIS SAMPLES

BAYESIAN NEURAL NETWORK OUTPUTS FOR THE tb+tqb ELECTRON ANALYSIS

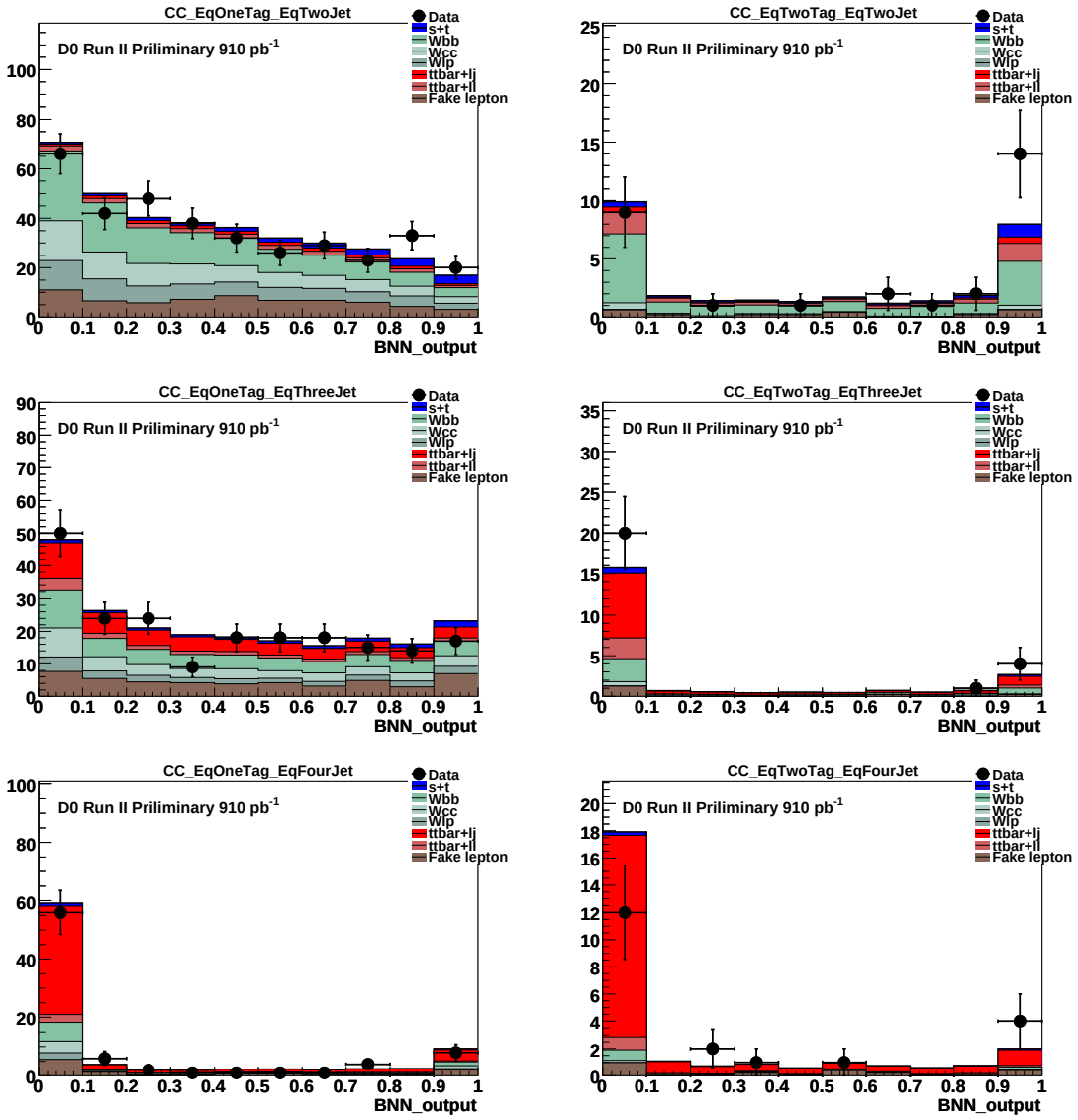

FIG. 18: Bayesian neural network output for the $tbtqb$ signal in the electron channel. Left column: one b-tagged jet; right column: two b-tagged jets. Top row: two jets; middle row: three jets; bottom row: four jets.

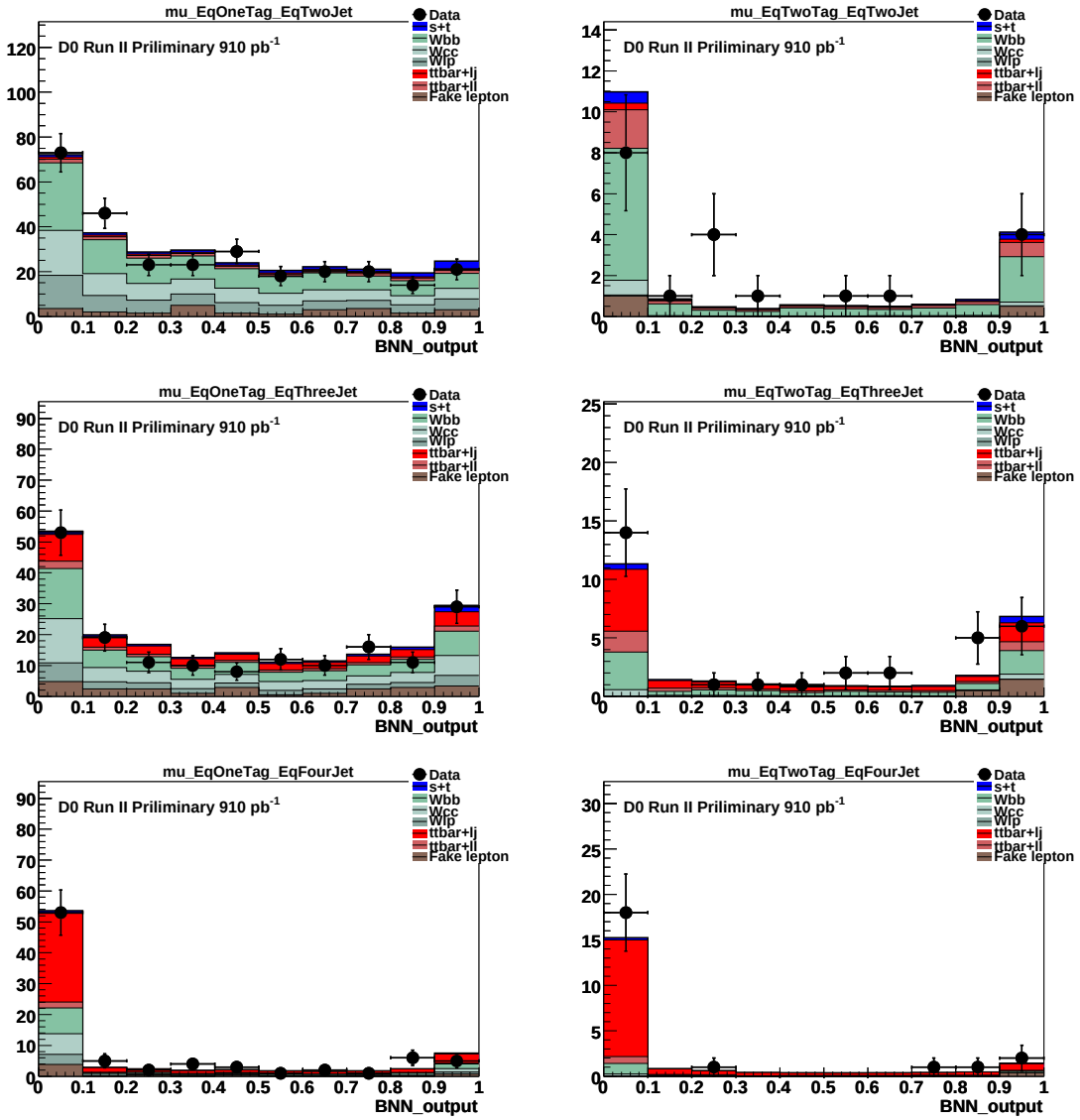

FIG. 19: Bayesian neural network output for the *tbtqb* signal in the muon channel. Left column: one *b*-tagged jet; right column: two *b*-tagged jets. Top row: two jets; middle row: three jets; bottom row: four jets.